DOCUMENT RESUME

ED 244 985 TM 840 307

AUTHOR Yap, Kim Onn
TITLE Standards for Title VII Evaluations: Accommodation for Reality Constraints.
PUB DATE Apr 84
NOTE 33p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Bilingual Education Programs; Elementary Secondary Education; *Evaluation Methods; Evaluation Utilization; Family Influence; *Program Evaluation; Program Implementation; Quality Control; School Community Relationship; *Standards; Test Reliability; Test Validity; Validity
IDENTIFIERS *Elementary Secondary Education Act Title VII; Multitrait Multimethod Techniques

ABSTRACT

Two separate sets of minimum standards designed to guide the evaluation of bilingual projects are proposed. The first set relates to the process in which the evaluation activities are conducted. They include: validity of assessment procedures, validity and reliability of evaluation instruments, representativeness of findings, use of procedures for minimizing error, and use of multiple objectives and multiple measures. The second set of standards relates to the content of the evaluation, and includes: project implementation; student performance; school, family, and community factors; and evaluation use. In implementing evaluation standards, several issues and problem areas are likely to emerge, e.g., resistance to change, burden on resources, and technical issues. However, the development and implementation of sound evaluation standards should go a long way in ensuring the accomplishment of desired outcomes in Title VII projects. (BW)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED244985

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. O. Yap

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

# Standards for Title VII Evaluations: Accommodation for Reality Constraints

Kim Onn Yap

Northwest Regional Educational Laboratory

A paper presented at the annual meeting of the
American Educational Research Association
New Orleans, April 23-27, 1984

Tm 820 307

# Standards for Title VII Evaluations: Accommodation for Reality Constraints

## INTRODUCTION

Much has been said about the lack of use of evaluation data by decisionmakers (Wise, 1978; Thompson and King, 1981; Berke, 1983). A host of factors have contributed to the use or non-use of evaluation information in compensatory education (Alkin et al., 1982). These include evaluator credibility, evaluator commitment to use, interest in evaluation by decisionmakers and the community, local focus of evaluation, effective presentation of results, assistance in developing procedures for the use of evaluation data. Title VII evaluations suffer additional obstacles created by the lack of technically sound and practical standards for the conduct of program evaluation. The present paper proposes several standards designed to guide the evaluation of bilingual projects. It is hoped that the implementation of these standards will not only substantially improve the technical adequacy of such evaluations but also enhance their potential usefulness to decisionmakers.

The development of the proposed standards is based on a review of the relevant literature (e.g., Bissell, 1979; Berke, 1980; Berke, 1983) and field experiences in using similar standards in other compensatory education programs (e.g., Chapter 1). There is ample evidence that bilingual education projects are among the most difficult to implement. A large degree of organizational change and mutual adaptation is required

to successfully implement a bilingual education project. Local capacity building and strong commitment supported by a well-planned inservice program are also needed. Evaluation of bilingual education programs faces many major obstacles, including limitations in existing instruments, problems in the use of comparison groups, contamination of the effects of school and community contexts, and the need to measure (not simply assume) project implementation.

These projects face additional problems of periodic refundings, uncertain renewals and program decisions beyond the control of project personnel. Moreover, parents, community members, program managers, school district, state and various federal decisionmakers may agree to only a few priorities for program implementation and evaluation.

Experience in implementing evaluation standards in other compensatory education projects indicates that what is most practical is what most often gets implemented in the local school setting. Thus, evaluation procedures which are superior in scientific rigor are often not used while less rigorous processes are put in place when the former are perceived to be esoteric or too complex. Impracticality is feared more than scientific invalidity. It is imperative that the development of evaluation standards takes into account real-life constraints which often dictates a compromise between scientific rigor and practicality.

To retain an appropriate level of flexibility, the proposed standards are not intended to be minimum acceptable levels which Title VII projects must achieve. Rather, they describe characteristics all Title VII projects must strive to attain. The extend to which a Title VII project

can meet these standards will be influenced by many factors including high transiency of the student population and limited availability of appropriate measurement instruments.

## PROCESS STANDARDS

Two separate sets of minimum standards are proposed. The first set relates to the process in which the evaluation activities are conducted and may be referred to as process standards. These standards include:

- o Validity of assessment procedures
- o Validity and reliability of evaluation instruments
- o Representativeness of findings
- o Use of procedures for minimizing error
- o Use of multiple objectives and multiple measures

### Validity of assessment procedures

This involves the use of experimental and quasi-experimental designs for conducting the evaluation, including the use of actual or statistical comparisons to show that a change did occur as a result of a Title VII project. This standard addresses such questions as (a) Did a change occur? (b) How likely is it that the observed effects resulted from the intervention? and (c) Is the presented evidence believable and interpretable? (Tallmadge, 1977).

While it is ideally desirable to implement a social intervention in a true experimental design, real-life constraints often dictates a compromise. Thus there exists a tension between scientific rigor and practicality of Title VII evaluation activities. If experience in implementing educational change efforts is any indication, what is most

practical is what is actually done in most, if not all, instances. For example, in spite of the relatively superior scientific rigor of the comparison group model and the regression model, the norm-referenced model in the Title I Evaluation and Reporting System (Tallmadge et al., 1981) is used by over 95 percent of the LEAs across the country. Thus, procedures which would produce the most valid assessment of project impact may often not be used because they are not the most feasible.

Moreover, in most cases, the use of a comparison group is either legally infeasible or is precluded by resource constraints. In such cases, statistical comparisons (e.g., local or national norms) would need to be used. The tradeoff is, again, between practicality and the ability to produce a strong causal link between effects and intervention. However, even in the absence of an actual comparison group the evaluation data can often be <u>suggestive</u> of project impact or can be used for program improvement purposes.

While misuse of statistical procedures (e.g., confidence tests) has been rampant in educational research and evaluation (Coats, 1970; Cronbach, 1975; Carver, 1978; Mook, 1983) a properly conducted test of significance remains a sine qua non for differentiating between random fluctuation and a reasonable estimate of program impact. Whenever feasible, such tests should be performed on Title VII evaluation results. In addition, when extreme subgroups (e.g., language dominance 1 students) are encountered in the evaluation, proper procedures should be used to avoid biased (inflated) estimate of program impact by reducing or eliminating regression effects (Thorndike, 1942; Campbell and Erlebacher, 1970; Campbell and Boruch, 1975; Bryk and Weisberg, 1977). Such procedures include using separate measures for selection and pretest (Tallmadge, et al., 1981).

With the above caveats, the following guidelines are proposed to ensure validity of assessment of program impact:

- Student performance should be assessed at at least two time points (e.g., pretest and posttest) to measure change in achievement status.
- Whenever feasible a comparison group (actual or statistical) should be used to measure achievement growth attributable to the project treatment.
- Some longitudinal followup assessment should be made of exited students to evaluate sustained effects of the intervention.
- An appropriate type of scores (i.e., those with an equal interval scale) should be used in assessing achievement gains.
- In cases where test norms are not available or not appropriate (e.g., in projects with severe problems of transiency or attrition) a criterion-reference approach may be used to conduct the evaluation.
- An attempt should be made to separate project effects from the effects of other school and community factors such as the implementation of other federal or special projects within the same schools. In cases where such contamination of effects cannot be ruled out, a statement should be made to point out that possibility.

Validity and reliability of evaluation instruments

The validity of an evaluation instrument is the extent to which it measures what it is intended to measure. For example, if the instrument does not measure what the Title VII project teaches, the results will not

provide useful or correct information about project impact. It is imperative that there be a good match between test items and the program curriculum or program objectives. In a valid instrument:

- o the items appear to match what the program teaches (content validity)
- o the development/selection of items is based on or is consistent with some theory (construct validity)
- o the results correlate highly with those obtained from similar instruments (concurrent validity)
- o very different scores are obtained for persons known to differ on the trait being measured (predictive validity)

The reliability of an instrument is an indication of how consistently it measures the trait it is intended to measure. In assessing achievement gains, for instance, a pretest and posttest are typically used to measure change in achievement status. If the test produces inconsistent results or if the results are affected by extraneous conditions, the change will be obscured. Test scores provided by a reliable instrument for a group of students will fall into about the same rank order (a) on two successive administrations of the instrument within a short interval, (b) on alternative forms of the same instrument and (c) when only the odd-numbered questions are scored as when only the even-numbered questions are scored. These are referred to as (a) test-retest reliability, (b) alternate-form reliability and (c) split-half reliability, respectively.

The standard of validity and reliability of instruments addresses the question of whether a change did occur and has a bearing on the

statistical and educational significance of the evaluation results (Tallmadge, 1977).

In selecting or evaluating a standardized achievement test, the standard of validity and reliability may be expanded to include the following criteria:

Measurement validity. This set of criteria looks at the nature of what a test measures, the range of behaviors sampled, the relationship of the test score to other measures, and the demonstrated usefulness of the test in theoretical or practical settings.

Examinee appropriateness. These criteria relate to the appropriateness of the test materials, including content of the stimuli (items) and mode of response, relative to the grade level of students taking the test.

Administrative usability. These criteria deal with practical concerns in administering and using a test. The ease with which the test can be given, scored, and interpreted, and the usefulness of the resulting score in making program or instructional decisions.

Technical excellence. These criteria are concerned with the test's reliability, replicability and refinement of measurement.

These criteria are described more fully in documents produced by the Center for the Study of Evaluation of UCLA (Hoepfner, et al., 1976), the Center for Bilingual Education (Silverman, et al., 1976; Silverman, et al., 1978) and the Assessment Projects at the Northwest Regional Education Laboratory (Nafziger, et al., 1975), the American Psychological Association, the American Educaional Research Association, the National

Council of Measurement in Education (Davis, et al., 1974), as well as individual researchers (e.g., Madaus, et al., 1982).

In using the above criteria it is imperative that input be obtained from project staff to help determine the instrument's validity and reliability within the context of the local project. In a recent study, Yap (1983) included perceptions of project staff as a criterion for test evaluation. The study showed that such an approach not only is feasible but also provides a consumer-oriented dimension to test evaluation.

Representativeness of findings

For the program manager to use Title VII evaluations, the results must be representative. That is, they must reflect as accurately as possbile the effects of the program on all students who participated in the program. The evaluator must decide whether to base the evaluation on all project students (i.e., the population ) or on a representative sample. The results obtained from a sample are representative if they do not differ systematically from those which would have been obtained had data been collected from the population.

Using a sample may significantly reduce the data collection burden on students and other project personnel as well as the amount of time needed to process and analyze the data. However, the sampling process requires a high level of technical expertise often not found in a local education agency. Relative advantages and disadvantages should be considered carefully before a decision is made on sampling. The evaluator's decision must satisfy the criterion that the evaluation results accurately reflect the effects of the Title VII project on its participants. Any sampling which precludes this should not be attempted.

Whether a sample or the population is used in the evaluation, it is more than likely that some students or classes will no longer be available for posttesting. The loss--referred to as attrition -- is most likely to occur in Title VII projects where the student populations are highly transient. If a large percentage of project students is not available for posttesting, the effects of attrition must be checked to ensure that the results are still representative. A lack of representativeness is indicated if:

- o the average score of students having both pretest and posttest scores differs significantly from the average score of students having pretest scores only, or
- o certain subgroups (e.g., those in language dominance category 1) have pretest scores but no posttest scores.

The standard of representativeness of findings addresses the question of whether the presented evidence is believable and interpretable and is a precursor to generalizability which addresses the question of whether the intervention can be implemented in another location with a reasonable expectation of comparable results. (Tallmadge, 1977).

Procedures for minimizing error

Title VII project evaluation plans should include procedures for minimizing error in (a) the administration of evaluation instruments, (b) scoring of instruments, (c) recording data, and (d) validating results.

- o Adminstration of instruments. The posttesting conditions and procedures must be consistent with the pretesting conditions and procedures.

- o Scoring instruments. In cases where instruments are scored by local project staff, at least a small sample of the measures should be scored independently by two individuals and the results compared to ensure comparability. In cases where instruments are scored by a commercial scoring service, results should be spot-checked for accuracy.
- o Recording data. Data recording forms should be designed to encourage accuracy and all data transcriptions should be proofread.
- o Validating results. At least a small sample of evaluation results should be recomputed to ensure correctness of computation.

The standard of error minimization addresses several critical questions presented in the Ideabook (Tallmadge, 1977), including (a) Did a change occur? (b) Was the effect consistent enough and observed often enough to be statistically significant? (c) Is the presented evidence believable and interpretable?

Multiple objectives and multiple measures

A Title VII evaluation typically serves many audience groups who have divergent information needs. It is important that multiple measures be used in the evaluation to address multiple program objectives. An evaluation with a narrow scope (e.g., summative achievement data) generally does not provide sufficient information for program managers to plan and carry out program improvement activities. Furthermore, the evaluation should be attentive not only to single program objectives but also to over-arching community objectives. For example, in a program

serving Indian student populations the maintenance of Native American languages and the transmission of cultural values may be an important objective to be addressed in the evaluation.

To the extent possible, a variety of methods (e.g., questionnaire, interview, document review, observation) should be used to collect evaluation information. The use of multiple strategies makes it possible to triangulate measures to achieve convergent validity (Cronbach, 1982; Odom and Fewell, 1983).

## CONTENT STANDARDS

The second set of standards relates to the content of the evaluation and may be referred to as content standards. These standards include:

- o Project implementation
- o Student performance
- o School, family and community factors
- o Evaluation use

### Project implementation

Program evaluations are often conducted without first ascertaining whether a program has been put in place. Such evaluations are potentially useless to decisionmakers. An assessment of program implementation is particularly important in bilingual education projects because these projects often face unique difficulties in program implementation (Bissell, 1979). Program managers frequently have to cope with problems such as insufficient numbers of adequately trained staff or an absence of appropriate materials and curricula. On the other hand, project activities that are potentially effective may not be fully implemented and may, therefore, appear to be ineffective.

An assessment of the degree and quality of program implementation allows the evaluator to analyze project impact in fully-implemented and partially-implemented sites. Information on degree of implementation allows the evaluator to make a more valid interpretation of project outcomes and is often of direct utility to the project staff. For example, information can be obtained in a program implementation evaluation on such matters as:

- The extent to which planned instructional approaches are used by the project staff
- How well the project staff have been trained to carry out the project activities
- The degree to which the instructional materials fit the performance level of project students

In addition, descriptive data on student characteristics, types of services provided, length of student participation and criteria for determining language proficiency are useful in project management. Program implementation information can also help identify key factors which influence the success of the project.

In evaluating project implementation, it is not sufficient (it is indeed irrelevant) to demonstrate that the adopted instructional procedure is different from others. Evidence must be obtained to show that the procedure has been implemented as intended (Shaver, 1983). Furthermore, project implementation information should have sufficient specificity to allow for the identification of effective program components for potential replication and dissemination.

In evaluating program implementation, the evaluator serves as a facilitator (Seidman, 1983), advisor (Alkin and Daillak, 1979), educator

(Cronbach, 1980) and negotiator/fact finder (Krathwohl, 1980). These roles appear to be most conducive to producing educational change. Assistance and trust, as opposed to coercion and distrust, are among the most effective ways to bring about program improvement (Siedman, 1983).

Student performance

The ultimate beneficiaries of a Title VII project are the bilingual students participating in the project. A critical element of program evaluation pertains to an assessment of the project's impact on student performance. Standardized test instruments and other assessment methods (e.g., interviews, questionnaires, observations, structured tasks, rating scales) can often be used to assess:

- language proficiency and dominance, and
- achievement in English and the primary language.

In assessing student performance it is imperative that the instructional validity of the assessment instrument be proven. In other words, participants should, by the end of the project, have received adequate instruction relevant to the tasks or skills which are tested (Fisher, 1983; Popham, 1983). Furthermore, the test must not be culturally or racially biased.

Whenever appropriate, the assessment of student performance should include non-cognitive areas such as affective and attitudinal changes as well as social skills. Projects serving bilingual students often have primary objectives in these areas and the attainment of these objectives should be measured.

School, family and community factors

Another essential element of the evaluation of bilingual education programs pertains to the influence of school, family and community

factors on project outcomes. School staff, parents and community members vary in their attitudes toward particular languages, in their support of bilingual education, and in their willingness to promote the use of languages other than English in the classroom (Bissell, 1979). It is important that the evaluation attends to roles of parents in programs, to the community at large, and to institutional contexts in which the program operates. Inclusion of these areas in the evaluation calls for the following activities:

- o Documenting the environment in which the program operates
- o Examining parent participation, including roles and functions of parent advisory councils
- o Determining the impact of the program on educational and other institutions within the community
- o Identifying effects of the program on families of participants, the primary language groups involved and the community at large

Evaluation use

The ultimate worth of an evaluation is measured by the extend to which the findings are used to make corrective actions for program improvement. In spite of widespread claims that evaluation is of little use for policymaking it has been increasingly evident that evaluation findings are used by policymakers (Caplan, et al., 1975; Rich, 1977; Weiss, 1977). In bilingual education, Berke (1983) showed that the AIR study (AIR, 1977, 1978), for example, has had a strong influence on both the Executive Branch and the Congress in formulating national policies on bilingual education. In other compensatory education projects, Alkin et al. (1982) reported that evaluation data were used at all decision levels by state and local education agencies.

The researchers found that different kinds of evaluation data had relative utility at the various organizational levels. School boards, district advisory committees and external agencies relied on summative data more extensively than other evaluation data. At the district administrative level, summative data were mixed about equally with other evaluation data developed by the district. At the building level, principals, coordinators and the like relied slightly more on project impact data than on other data. At the classroom level, impact data were less often used. Instead, data more closely related to the instructional programs were preferred. Analysis of case studies showed that evaluation use was affected by several contextual variables, including:

Evaluator credibility. The reputation and credibility of the evaluator is an important determinant of use. While evaluators may achieve credibility in differing ways they must be perceived as competent and trustworthy.

Evaluator commitment to use. Credibility, while important, is not enough to insure evaluation use. The evaluator must also have a commitment to seeing that evaluation results are used by decision makers.

Interest in evaluation by decisionmakers and the community. Evaluation data are used when they are tailored to the needs and interests of the local school community. Use occurs when evaluators draw relevant information from evaluation data and when they conduct special evaluations to meet local requests.

Local focus of evaluation. Use increases when evaluations are specifically designed to meet local needs. Success of use is attributable to timely response and sensitivity to local concerns.

Effective presentation of results. Graphic, narrative and nontechnical modes of presentation increase the utilization of evaluation data by local decisionmakers.

Assistance in developing procedures for the use of evaluation data. Evaluation use increases when decisionmakers are assisted in understanding how they might use the evaluation data. Successful evaluators typically provide detailed, step-by-step procedures to potential users.

Alkin et al. (1982) suggested that state and local evaluation units should be encouraged to design a variety of local decision-focused evaluation strategies. In particular, locally designed evaluation procedures might provide information on the impact and costs of various materials and processes within projects. The researchers pointed out that many local and state agency personnel required guidance in developing procedures to follow when making decisions. It was not that administrators did not want to use relevant information. They typically did not know how to incorporate the information into their decision processes. Several steps can be taken both during and at the completion of an evaluation to increase the likelihood of its use:

- Mechanisms are developed for obtaining staff reactions to evaluation findings and recommendations
- Project staff are involved in identifying and analyzing potential corrective actions to address evaluation findings
- Project plans are revised periodically to include corrective actions
- Specific strategies are developed to implement the corrective actions

- Followup procedures are developed to evaluate progress in implementing corrective actions

## IMPLEMENTING THE STANDARDS

The process and content standards can be used as guidelines for implementing all Title VII evaluation activities, including:

- Planning and organizing for the evaluation
- Designing the evaluation
- Measuring project implementation
- Measuring student performance
- Measuring family, school and community factors
- Analyzing and reporting results
- Using evaluation findings

As indicated earlier, the standards are not intended to be absolute requirements with which Title VII projects must comply. They should, instead, be used as ideals to which a Title VII evaluation must approach. The adequacy of the evaluation is measured by the closeness with which it comes to meeting the standards. Title VII evaluators are faced with a growing schism between academe and practice. The range of skills and temperament required for each are different, ranging from precision and methodological sophistication in the case of research analyses to the more pragmatic, decision-oriented approach to program implementation and evaluation. In some cases this may lead to a tension between the quest for scientific rigor and technical excellence on the one hand, and the desire to provide responsive, timely and effective help in reaching decisions on the other. The perspective of project staff is undoubtly also influenced by current debates on whether the dominant

realities of evaluation are political or technical. In the former point of view, evaluation is an intimate part of the political process and its success will be partly political (Pincus, 1980). In the latter viewpoint, methodological and communications improvements will lead to success (Boruch and Cordray, 1980). In describing the widening gulf between academe and the real world, Stanfield (1981) says: "The academic view of the subject is pure, exact, permitting sophisticated methodologies in simplified and abstracted settings. The real world is pragmatic, oriented towards useful results rather than theoretical purity, and constrained by time and cost." In this regard it is important to realize that project staff are primarily concerned with what is "doable" in the local district setting rather than what constitutes the ideal. Furthermore, they are primarily concerned with the well-being of project participants rather than the advancement of knowledge. They serve first as providers of instruction and secondarily as promoters of science and knowledge.

In implementing the standards, care should be taken to ensure that the standards are compatible with both federal regulations and state policies where such policies exist. Title VII staff should review and update the standards periodically. The standards should be revised on the basis of knowledge and experience gained during implementation. Procedures and practices which are:

(a) not meaningful or useful to state or local education agencies,

(b) impractical to implement, or

(c) inconsistent with federal regulations and state or local policies

should be eliminated or modified and improved. It is expected that a final set of standards that is both practical and technically sound in a particular local context will emerge from this evolving process.

To facilitate standards implementation, a set of standard forms may be developed for the collection, analysis and reporting of data to the various levels of educational agencies. The forms will provide for the collection and aggregation of data from the building level upward through the local and state education agencies.

The forms may include information such as project description, project implementation and student achievement. Project description and implementation information may include instructional objectives, number of participants, ethnic backgrounds of participants, project duration, project setting, instructional approach, teacher-student ratio, class size, project funding level, per pupil cost, parent advisory council activities, total hours of instruction, hours of instruction per week, inservice training for project staff including topics, number and duration of training sessions.

Student achievement information may include pre-project and post-project achievement status, achievement gains and/or percent of participants attaining specified instructional objectives. Achievement information should be documented by grade level. Where achievement data are aggregated across school buildings and projects, weighted averages should be used.

## PROBLEMS AND SOLUTIONS

In implementing the evaluation standards several issues and problem areas are likely to emerge. These potential problems and their proposed solutions are discussed below.

Resistance to change

If the proposed use of the standards is seen as an external force attempting to impose change on the local or state education agencies, strong resistance may manifest itself in many subtle and disguised ways (Insel and Moos, 1974). Such manifestations range from legitimate questioning of the technical adequacy and usefulness of the standards to a perfunctory implementation to satisfy compliance requirements. The standards may appear to some to be an attempt to usurp local prerogatives by prescribing program evaluation practices to states and local districts. Some initial resistance to the implementation of the standard is to be expected. Such resistance could, if not deftly dealt with, greatly reduce the usefulness of the standards.

Most important to overcoming resistance will be the evaluator's success in establishing credibility with project staff. Evaluators should be selected in part for their strength in interpersonal skills and communications.

Specific solution strategies for reducing resistance include:

- Providing materials designed to (a) increase awareness of the value of improved practices in program evaluation, and (b) explain in lay terms the ways in which the use of the standards can be helpful in specific situations
- Providing services which complement functions and responsibilities of Title VII project staff
- Conducting needs sensing activities to ensure that project staff's needs are met through the provision of technical assistance

Burden on resources

Implementation of the standards may, in some cases, result in an increased need for human and fiscal resources at the state and local levels. Furthermore, it will, in most cases, demand increased technical capability among project staff responsible for program management, documentation and evaluation. It is likely that as the standards are implemented, a reallocation of project resources and priorities will occur. In some cases, such reallocation may result in reduction of classroom services. Some educators will view this outcome as undesirable and may object to the trade between improved evaluation and reduction in student services.

The only real justification for evaluation is that it leads utimately to improved programs and services for children. At the state level, improvement might mean developing capacity for providing meaningful and valuable advice and counsel to local school districts about successful program practices. At the local level, improvement might mean making program changes because evaluation data showed that changes are needed.

Evaluators should focus their work on improving program practices so as to improve educational opportunities for children. Two primary tasks will be:

- To alter attitudes towards evaluation by demonstrating its worth as one means of improving educational opportunities
- To provide the type of assistance to upgrade management, documentation and evaluation practices for positive impact on student services

Technical issues

Several technical issues will arise in implementing the minimum standards. These include (a) problems stemming from the transient nature of the student population, (b) contamination of evaluation results, (c) flexibility in evaluation procedures and (d) divergent information needs of different audience groups. Each of these issues is discussed in further detail below.

Transiency. Title VII projects serve a relatively transient student population. The "exit" rate in some cases may result in a very small number of project students being included in the evaluation. This attrition problem poses a severe threat to the representativeness of the evaluation findings. Strategies for resolving this problem include (a) use of tests with monthly or quarterly norms which permit more students to be pre- and posttested regardless of their length of stay in the project, (b) use of criterion-referenced measures which permit students to be tested as they enter and leave the project, and (c) use of separate comparison standards for subgroups of project students based on length of time spent in the project.

Contamination of result. State and local education agencies may receive multiple sources of special funding from the federal government. This could result in more than one project treatment being provided to the same student population. In such cases, outcomes of one project are confounded with effects of other treatments provided to the same student groups. Although it is possible in some cases to disentangle the effects of multiple program implementation on student performance, most projects -- especially those in the smaller districts -- are not likely to have the

resources or staff expertise to undertake a highly sophisticated evaluation study. Specific strategies for addressing the contamination of results will include (a) identifying all special program services provided to the project students, (b) separating the effects of different projects whenever feasible, and (c) acknowledgingthe contamination if separation of effects is not feasible.

It should be noted that even when the separation of effects is not feasible, the evaluation data may still be suggestive of program impact (or the lack of it) and often are useful for program improvement purposes. Furthermore, in districts where multiple sources of special funding exist, project managers may consider collaboration with the other funding sources in conducting program evaluation.

Flexibility in evaluation procedures. The implementation of the minimum standards provides for a great deal of flexibility on the part of project staff in using these standards. Some may have been accustomed to complying with a specific set of rules and requirements and, as a result, are less comfortable when presented with the more flexible minimum standards.

Instead of being pleased with the number of options available to them in using the standards, project staff with limited evaluation skills may be disappointed that specific procedures were not prescribed. This attitude may be more prevalent in medium and small districts where personnel assigned to evaluation tasks have numerous other responsibilities. With more options available, project staff will need additional assistance in understanding and selecting from a

range of options. Evaluators can lessen the burden by means of the following strategies:

- o Options for using the minimum standards that are most compatible with existing practices should be emphasized.
- o Benefits of recommended procedures and practices congruent with local needs should be highlighted.
- o Project staff should be encouraged to use the minimum standards to collect information most useful for decisionmaking and program improvement.
- o Assistance should be provided to increase project staff's awareness of factors affecting evaluation use so that relevance of evaluation in meeting local needs is emphasized.

Divergent information needs. It is recognized that different needs for evaluation information exist among the various levels of educational agencies involved. At the project level, data must be responsive to the needs of teachers, parent advisory councils, project managers and district administrators. At the state level, data on student performance need to be summarized across projects and compiled for the state as a whole. State board of education priorities and legislative reporting requirements must also be attended to. The U.S. Education Department needs data which can be aggregated for many different projects for reporting to Congress. These divergent information needs will compete for limited resources available to the state and local education agencies.

Specific strategies for resolving the problem include the following:

- o Standard data collection, analysis and reporting forms provide a partial solution to the problem. These forms establish a common

data base at each local education agency and provide for the collection and aggregation of data from building level upward through the state to the U.S. Department of Education.

- Evaluators should provide a rationale to project staff for the proper use of the forms and help them understand the possibilities of using a common data base to supply information for multiple audiences and develop the capacity to generate such a data base.
- If a local education agency is faced with resource constraints, evaluators should provide the project staff with training in prioritizing information needs and in using appropriate criteria for selecting evaluation questions to be addressed.

## CONCLUDING REMARKS

The implementation of evaluation standards is expected to bring about improved quality of evaluation data and increased use of such data for program improvement. The movement toward effective schooling has been gathering momentum during the past several years. With the movement reaching full swing, it is not surprising that evaluation activities will be stepped up as a means of achieving accountability and quality control of local, state and federal efforts in education. The development and implementation of sound evaluation standards will go a long way in ensuring the accomplishment of desired outcomes in Title VII projects.

That this is both doable and desirable is demonstrated by recent efforts in implementing a set of federally initiated evaluation procedures in local Title I/Chapter 1 projects (Stonehill and Anderson,

1982). Through a program of technical assistance and a process of mutual adaptation and refinement, the concept was found to be "working and working well" within reality constraints (Millman, et al., 1979; Yap, 1983). There is evidence that state and local educators and evaluators working with Chapter 1 projects are now more knowledgeable about issues in educational evaluation than they were prior to the implementation of the evaluation standards (Reisner, et al., 1982). The implementation of standards has resulted in improvements in many areas including program improvement evaluation, testing procedures, needs assessment, quality control systems, program sustained effects and the identification of exemplary projects (Stonehill and Anderson, 1982). There appear no reasons why similar improvements cannot be made in Title VII projects.

## REFERENCES


Alkin, M.C., and Daillak, R.H. A study of evaluation utilization. Educational Evaluation and Policy Analysis, 1979, 1(4), 41-49.

Alkin, M.C., Stecher, B.M., and Geiger, F.L. Title I evaluation: Utility and factors influencing use. Washington, D.C.: U.S. Department of Education, 1982.

American Institutes for Research. Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program (Vol. 1). Palo Alto, Calif.: AIR, 1977.

American Institutes for Research. Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program (Vol. 4). Palo Alto, Calif: AIR, 1978.

Berke, I.P. Evaluation into policy: Bilingual education, 1978. Unpublished dissertation, Stanford University, 1980.

Berke, I.P. Evaluation and incrementalism: The AIR report and ESEA Title VII. Educational Evaluation and Policy Analysis, 1983, 5(2), 249-256.

Bissell, J.S. Program impact evaluations: An introduction for managers of Title VII projects. Los Alamitos, CA: Southwest Regional Laboratory for Educational Research and Development, 1979.

Boruch, R.F., and Cordray, D.S. An appraisal of educational program evaluations: Federal, state and local agencies. Evanston Illinois: Northwestern University, 1980.

Bryk, A.S., and Weisberg, H.I. Use of the nonequivalent control group design when subjects are growing. Psychological Bulletin, 1977, 84, 950-962.

Campbell, D.T., and Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Compensatory education: A national debate, (Vol 3). New York: Brunner/Mazel, 1970.

Campbell, D.T., and Boruch, R.F. Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In A. Lumsdaine and C. Bennet (Eds.), Experiment and evaluation. New York: Academic Press, 1975.

Caplan, N., Morrison, A., and Stambaugh, R.J. The use of social science knowledge in policy decisions at the national level. Ann Arbor, Mich.: CRUSK, Institute for Social Research, 1975.

Carver, R.P. The case against statistical significance testing. Harvard Educational Review, 1978, 48(3), 378-399.

Coats, W. A case against the normal use of inferential statistical models in educational research. Educational Researcher, June 1970. Vol. XXI, 6-7.

Cronbach, L.J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.

Cronbach, L.J., and associates. Toward Reform of Program Evaluation. San Francisco, Calif: Jossey-Bass, 1980.

Cronbach, L.J. Designing evaluation of educational and social programs. San Francisco: Jossey-Bass, 1982.

Davis, F.B. (Chair) Standards for educational and psychological tests. Washington, D.C.: American Psychological Association, 1974.

Fisher, T.H. Implementing an instructional validity study of the Florida high school graduation test. Educational Measurement Issues and Practice, 1983, 2(4), 8-9.

Hoepfner, R., Bastone, M., Ogilvie, V., Hunter, R., Sparta, S., Grothe, C.R., Shani, E., Hufano, L., Goldstein, E., Williams, R., and Smith, K.O. CSE elementary school test evaluations. Los Angeles: Center for the Study of Evaluation, UCLA, 1976.

Insel, P.M., and Moos, R.H. Psychological environment: Expanding the scope of human ecology. American Psychologist, 1974, 29(3), 1979-88.

Krathwohl, D.R. The evaluator as negotiations facilitator - fact finder. Educational Evaluation and Policy Analysis, 1980, 2(2), 25-34.

Madaus, G.F., Airasian, P.W., Hambleton, R.K., Consalvo, R.W., and Orlandi, L.R. Development and application of criteria for screening commercial, standardized tests. Educational Evaluation and Policy Analysis, 1982, 4(3), 401-415.

Millman, J., Paisley, W., Rogers, W.T., Sanders, J.R., and Womer, F.B. Performance review of USOE's ESEA Title I evaluation technical assistance program. Washington, D.C.: Hope Associates, 1979.

Mook, D.G. In defense of external invalidity. American Psychologist, 1983, 38(4), 379-387.

Nafziger, D.A., Thompson, R.B., Hiscox M.D., and Owen, T.R. Tests of functional adult literacy: An evaluation of currently available instruments. Portland, OR: Northwest Regional Educational Laboratory, June 1975.

Odom, S.L., and Fewell, R.R. Program evaluation in early childhood special education: A meta-evaluation. Educational Evaluation and Policy Analysis, 1983, 5(4), 445-460.

Pincus, J. (Ed.). Educational evaluation in the public policy setting. Santa Monica, California: The Rand Corporation, 1980.

Popham, W.J. Task-teaching versus test-teaching. Educational Measurement Issues and Practice, 1983, 2(4), 10-11.

Reisner, E.R., Alkin, M.C., Boruch, R.F., Linn R.L., and Millman, J. Assessment of the Title I evaluation and reporting system. Washington, D.C.: U.S. Department of Education, 1982.

Rich, R.F. Uses of social science information by federal bureaucrats: Knowledge for action vs. knowledge for understanding. In C.H. Weiss (Ed.), Using social research in public policy making. Lexington, Mass.: D.C. Heath, 1977.

Seidman, W.H. Goal ambiguity and organizational decoupling: The failure of "rational systems" progam implementation. Educational Evaluation and Policy Analysis, 1983, 5(4), 399-413.

Shaver, J.P. The verification of independent variables in teaching methods research. Educational Researcher, 1983, 12(8), 3-9.

Silverman, R., Noa, J.K., and Russell, R.H. Oral language tests for bilingual students: An evaluation of language dominance and proficiency instruments. Portland, OR: Northwest Regional Education Laboratory, July 1976.

Silverman, R., and Tupper, N. Assessment instruments in bilingual education. Portland, OR: Northwest Regional Educatonal Laboratory, 1978.

Stanfield, J. Management review of evaluation practice. Northwest Regional Educational Laboratory, Portland, Oregon: 1982.

Stonehill, R.M., and Anderson, J.I. An evaluation of ESEA Title I -- Program operations and educational effects. Washington, D.C.: U.S. Department of Education, 1982.

Tallmadge, G.K. Ideabook. Washington, D.C.: U.S. Department of Health, Education and Welfare, 1977.

Tallmadge, G.K., Wood, C.T., and Gamel, N.N. User's guide ESEA Title I evaluation and reporting system. Mountain View, CA: RMC Research Corporation, 1981.

Thompson, B., and King, J.A. Evaluation utilization: A literature review and research agenda. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

Thorndike, R.L. Regression fallacies in the matched groups experiment. Psychometrika, 1942, 7, 85-102.

Weiss, C.H. Research for policy's sake: The enlightenment function of social science research. Policy Analysis, 1977, 3(4), 531-546.

Wise, R. What we know about the decision maker in decision settings. Paper presented at the annual meeting of the Amercian Educational Research Association, Toronto, March 1978.

Yap, K.O. TAC serendipity: Random thoughts on unanticipated outcomes. A paper presented at the annual meeting of the American Educational Research Association, Montreal, 1983.

Yap, K.O. Evaluating a bilingual test: Adding the consumer's point of view. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1984.